DOCUMENT RESUME

ED 189 115 TM 800 278

AUTHOR Huntley, Renee M.: Plake, Barbara S.
TITLE Effects of Selected Item-Writing Practices on Test Performance: Can Relevant Grammatical Clues Result in Flawed Items?
PUB DATE Apr 80
NOTE 34p.: Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Entrance Examinations: *Cues: Difficulty Level: *Grammar: Higher Education: High Schools: *Multiple Choice Tests: Racial Differences: *Response Style (Tests): Sex Differences: Social Studies: *Test Construction: *Test Items: Test Results
IDENTIFIERS Act Assessment: *Distractors (Tests): Item Discrimination (Tests)

ABSTRACT

Guidelines for test item-writing have traditionally recommended making the correct answer of a multiple-choice item grammatically consistent with its stem. To investigate the effects of adhering to this practice, certain item formats were designed to determine whether the practice of providing relevant grammatical clues, in itself, created cue sensitivity both for misdirection and direction. American College Test (Act) Assessment experimental social science items were constructed to investigate grammatical compliance in two particular situations: plural-singular and vowel-consonant (stem ending in a/an) agreement. Items were worded so that all situations were grammatically appropriate. Results suggested that subjects were sensitive to these cues, to some extent. It was suggested that rephrasing items into neutral, but still grammatically precise language--for example, including words such as a/an, or agency/agencies in stems--could eliminate this sensitivity. (Eight illustrative items are appended). (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED189115

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Effect of Selected Item-Writing Practices on Test Performance: Can Relevant Grammatical Clues Result in Flawed Items?

Renee M. Huntley

American College Testing Program

Barbara S. Plake

University of Nebraska-Lincoln

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R. Huntley

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the 1980 convention of the American Educational Research Association, Boston, Massachusetts.

TM 800278

Effect of Selected Item-Writing Practices on Test Performance: Can Relevant Grammatical Clues Result in Flawed Items?

Abstract

This study examined the effect of adhering to the universally endorsed item-writing practice of making the correct answer of a multiple-choice item grammatically consistent with the stem. Formats were designed to investigate whether the practice of providing relevant grammatical clues in itself created cue sensitivity both for direction and misdirection. ACT Assessment experimental social science items were constructed to investigate grammatical compliance in two particular situations: plural-singular and vowel-consonant agreement. In no instance were grammatically inappropriate situations created. Results suggest some subject sensitivity to these cues. Rephrasing into neutral, but still precise and grammatically correct, language can eliminate this sensitivity.

Effect of Selected Item-Writing Practices on Test Performance: Can Relevant Grammatical Clues Result in Flawed Items?

Measurement texts warn test developers to avoid specific flaws in writing test items because flaws influence test scores (Ebel, 1972). Among the flaws suggested to affect test scores are window dressing, length of correct answer, completeness of item, irrelevant cues, and grammatical inconsistencies between item stem and incorrect alternatives. Several studies have investigated item-writing practices, and the results are mixed. Dunn and Goldstein (1959), McMorris, Snyder, Brown, and Pruzek (1972) found that flaws in alternative length and grammar produced easier items, while Board and Whitney (1972) failed to find significant effects.

Two types of grammatical agreement were investigated in this study: plural-singular and vowel-consonant. One strategy for ensuring agreement between stem and alternatives is to "generalize" the stem: plural verbs or vowel-matching articles are added, usually parenthetically, to provide for grammatical agreement with all options. Since the correct answer is typically the one that differs grammatically from the other alternatives in the set, thus necessitating the "generalizing" of the stem, a source of item cuing may be created that will affect performance.

For this study, four versions of each item were developed and labeled A, B, C, and D. Version A items contained a generalizing grammatical cue that was consistent with the correct alternative; each version B item had the same correct answer as its version A counterpart, but the item was rewritten so that a generalizing cue was omitted from the stem. Thus, versions

A and B helped investigate the function of generalizing cues for systematically directing students to the correct answer. Version C items included a generalizing grammatical cue in the stem, but agreement with the correct answer did not rely upon it; however, also included in the set of incorrect alternatives was one that was grammatically consistent with the generalizing language in the stem. Version D items had the same set of alternatives, but the stem was rewritten into neutral language, thus eliminating the grammatical cue. Therefore, versions C and D investigated the misdirection of students due to grammatical cues in the stem. (Examples of these items can be found in Appendix 1.)

A total of 16 items from a social science exam were used: 9 were designed to measure plural-singular agreement, and 7 were designed to measure vowel-consonant agreement. Items that measure plural-singular agreement were unique in that they were based on a "combination" format. The combination item format represents a logical extension to the typical specific-answer multiple-choice item. The alternatives group the options (usually identified as I, II, III) in specific combinations. The student confronts a more complex task since the most comprehensively correct combination of options must be selected. No research on the combination format was found, although items of this nature appear in many standardized tests. Inclusive or exclusive options ("all ... " or "none of the above") increase item difficulty (Dudycha & Carpenter, 1973; Hughes & Thumber, 1956), but whether this generalizes to combination items is yet to be studied.

The items constituted an experimental social science test administered as part of the ACT Assessment tryout. The items were randomized within the tryout units in such a way that each student had one of the four versions of each item, and all the item types were represented in each form.

Method

Four experimental social science units were administered to a total of 1051 students. The distribution of subjects across the four units, broken down by sex and race, can be found in Table 1. Item characteristics (difficulty and discrimination indices) were calculated for the items within the experimental unit in which they appeared, and the investigation of the effects of grammatical cues concentrated on these item characteristics. The items were then sorted into the appropriate research form (A, B, C, or D) for subsequent analyses.

Analysis

Overview

In order to analyze different aspects of cue sensitivity, four distinct analyses were selected. A general description of each analysis is provided first, after which the results are reported (plural-singular and vowel-consonant item types were analyzed separately; results for each are reported within each analysis performed).

The procedure used in developing the items for the four research forms consisted of systematically rewriting the same item to match the specifics of each form. The same item content, therefore, was varied across the forms. For plural-singular item types, 9 items were written four times; 7 vowel-consonant item types were also systematically varied in the same manner. Since the procedure systematically varied the presentation of the same item in four ways, an analysis of trends in performance across the forms was chosen. A three factor ANOVA design, with form considered a within factor, group and items perceived as a between factor, was selected to analyze trends in the percent of students choosing the correct alternative across the 4 alterna-

tive versions. In one analyses, sex (male, female) was the group factor; race (white, other) became the group factor in the second.

A three factor fixed effects ANOVA was selected to analyze the function of generalizing cues for plural-singular and vowel-consonant types. The three factors were status of cue (present, absent), direction of cue (leads, misleads), and group (separately analyzed for sex -- male, female -- and race -- white, other).

Whether the presence or absence of the generalizing cue directs students to the keyed response or misdirects them to an incorrect alternative was investigated by a separate series of analyses. Direction of the student by a generalizing cue in the item was investigated by singling out Form A items. The pattern of responses across the five alternatives was compared to a random pattern by means of a X^2 Goodness of Fit for each item type (plural-singular and vowel-consonant). Misdirection was analyzed by concentrating on Form C items only and comparing the response pattern across the incorrect alternatives with a random pattern, using a X^2 Goodness of Fit test. This analysis was performed separately for plural-singular and vowel-consonant items. The potency of the presence of the cues in the forms to direct or misdirect student responses was investigated by considering the popularity of consistent alternatives in both situations: when the cue in the stem is present or absent. For direction potency, Forms A and B were considered, and the proportion of students selecting the consistent alternatives was compared by calculating the average shift to the consistent alternative(s) for each of the item types considered. Misdirection potency was calculated in a similar fashion, by determining the average shift in popularity for the consistent, yet incorrect, alternative(s) from Form C to D.

Results

Form Analysis

Table 2 presents the means and standard deviations of percent of examinees, by sex and by race, choosing the correct alternative for the two item types. Results of the mixed model ANOVA (Form, within; Item, between; Group, sex or race, between) are summarized in Table 3. For plural-singular items, both the sex and race analyses found significant Group differences ($\bar{X}_M$ = 34.39, $\bar{X}_F$ = 38.87; $\bar{X}_W$ = 33.63, $\bar{X}_O$ = 27.09). Item and Item by Form effects were also significant, effects that are not of interest to this study. Group effects were also noted for vowel-consonant items ($\bar{X}_M$ = 52.43, $\bar{X}_F$ = 45.03; $\bar{X}_W$ = 50.11, $\bar{X}_O$ = 39.98), as well as significant Item and Item X Form effects. In addition, a significant Form X Group interaction was detected for the sex data. Figure 1 displays the means for each form by sex. A Bonferroni type (α = .05) follow-up was used to explain the interaction further. These results are shown in Table 4. Although males outperformed females on each form of the vowel-consonant items, a major discrepancy was found for Form C. Females performed more poorly in this form than in any of the others, a situation that suggests that females may be more sensitive to the subtle cuing present in the generalizing vowel-consonant items; because of this sensitivity, it would seem, they choose the correct answer less frequently. On all the forms of the tests, males outperformed females, and whites outperformed other racial groups.

Cue Analysis

The influence of the cuing on percent of examinees selecting the keyed alternatives was investigated by a three factor fixed effect analysis of variance [direction: lead, mislead; status: present, absent; group: sex

(male, female), race (white, other)] for both plural and vowel item types. A summary of the ANOVA results is shown in Table 5. For plural-singular type items, a significant race effect was found ($\bar{X}_W$ = 33.69, $\bar{X}_O$ = 27.09). No other significant effects were noted for these items. Sex and race were both significant for vowel-consonant type items ($\bar{X}_M$ = 52.43, $\bar{X}_F$ = 45.03, $\bar{X}_W$ = 50.11, $\bar{X}_O$ = 37.55). Another series of ANOVAs was processed using an index of discrimination (biserial correlation) as the dependent measure. Table 6 contains the means and standard deviations of the biserial coefficient. Table 7 displays the ANOVA summary tables for the analyses of item discrimination for both item types. Group was not considered as a factor in this design. No significant effects were found.

Cue Sensitivity

Direction

The presence of the subtle cue in the stem in Form A, if followed systematically by the examinees, would lead them to a set of consistent alternatives. If the students were sensitive to the cue, the proportion of students selecting consistent alternatives would be reflected in the pattern of responses across the alternatives. This pattern of responses was compared to a random pattern for both types of items. The results for plural-singular item types are shown in Table 8; they do not suggest a systematic pattern atypical from random. However, the results for vowel-consonant item types, which are shown in Table 9, do suggest a pattern of responses different from random.

The fact that students select the correct answer in an item that is possibly flawed due to the presence of a subtle cue does not necessarily

mean that the students were sensitive to the cue, following it to the keyed alternative. They may simply have selected the keyed answer because they were able to identify it as the best answer within the set of alternatives. In order to explain further the effects of a generalizing cue in directing students to the correct answer, the proportions of students selecting the "consistent" alternatives (i.e., alternatives that would be selected due to cue sensitivity alone) from Form A (where the cue is present) to Form B (where the cue is absent) were compared. The results of this comparison are shown in Table 10 for plural-singular item types. On the average, 1.4% more of the students selected the consistent alternatives when the cue was absent (Form B) than when it was present (Form A). The results of the comparison of the popularity of the "consistent" alternative for the vowel-consonant item types (when the cue is present or absent) are shown in Table 11. Across the seven vowel-consonant items, on the average, 0.7% more students selected the keyed alternative when the cue was absent. Therefore, even though it was found in the X^2 Goodness of Fit analysis that the pattern of responses for vowel-consonant item types was not consistent with a random pattern, the popularity of the keyed responses does not appear to be directly attributable to the presence or absence of the cue.

Misdirection

For each item in Form C, and for the students who missed the item, the percent who chose each of the four distractors was calculated. These percentages were summed separately for the nine plural-singular and seven vowel-consonant item types. An average was calculated that represented the percent choosing each incorrect alternative. Plural-singular item types, in every case, contained two alternatives that were consistent with

the subtle cue present in the item, choices D and E. If students were sensitive to the cue and allowed it to influence their choice of correct answer, half would have selected D and the other half E. In the absence of direction from the stem, the students who missed the item should, on the average, have selected each alternative an equal proportion of the time. Therefore, the distribution should have been 25% for each of the four incorrect alternatives. This model was used, and a X^2 Goodness of Fit test was performed to test the observed distribution of choices across the four incorrect alternatives and compared to a distribution that would be expected in the absence of cuing. A summary of the results is shown in Table 12. The X^2 value for the data was 5.04, which was not significant since $_{.95}X^2_3 = 7.82$. Therefore, in general, the pattern of selection of alternatives does not reveal one systematically different from random for plural-singular type items.

For the vowel-consonant item types, only one alternative is consistent with the presence of a generalizing cue, the alternative that begins with a vowel. The location of that alternative in the A to E placement varied randomly within the set of items. For the analysis of the distribution of incorrect choices, the location of the consistent alternative, which was actually random in the items, was changed to location E. For each item, the distribution of percent of incorrect choices was determined and averaged for consistent and other alternatives. This distribution was compared to a random pattern using a X^2 Goodness of Fit test. The results are shown in Table 13. The X^2 value, 5.24, did not surpass the $_{.95}X^2_3$

critical value of 7.82. Therefore, as was found with the plural-singular type items, the pattern of selection of incorrect alternatives does not support a systematic influence of the generalizing cue.

There is another way to view the influence of the cuing process: it may be identified by a shift in the popularity of an alternative that was consistent with the cue, given the presence or absence of the cue. Form C items contain a generalizing cue that, if followed by the examinee, would lead to the selection of a consistent alternative. When the cue is removed or "neutralized," a shift in the number of students selecting the consistent alternative would be noted. Therefore, if students were responding to the cuing situation created by the generalizing language of Form C, the neutralizing of the cue in Form D should have resulted in a shift away from the consistent alternative, as evidenced by fewer students choosing the same alternative in Form D.

Plural-singular type items in Form C contain two alternatives that are consistent with the generalizing cue present in the stem. The proportion of students choosing either was summed for each of those nine items. The same record was made of the same alternatives for those items in Form D. The difference by item was calculated from the proportion choosing alternatives consistent from Form C to Form D. Table 14 contains the proportions selecting consistent alternatives from Form C to D. The average change in this proportion was .018 in the direction of the cue, a negligible shift.

For the seven vowel-consonant type items, a similar computation was made that noted and then subtracted the proportions of students selecting

the consistent alternatives in Form C and Form D to obtain the difference in the proportions selecting the consistent alternative from Form C to Form D. The results of the computations are shown in Table 15. The average shift from Forms C to D was -.016, a negligible shift.

Conclusions

In general, the amount of cuing produced by stating stems in a "generalizing" fashion appears to be minimal. Total performance, as measured by percent choosing the correct answer, does not change significantly regardless of the status of the generalizing cue. However, women may be more acutely sensitive to cues present in vowel-consonant items, since women performed differentially less well on the form that combined cue presence and misdirection (Form C).

It is interesting to note that plural-singular type items were consistently more difficult across experimental forms than were vowel-consonant items. This finding confirms the expectation that the plural-singular item type which uses the combination format presents a more complex question to the student. More research is needed to investigate the combinational item type.

The consistent superiority of males across test forms and item types was surprising since social studies items were used. The performance of presumably more verbally talented females is traditionally expected to be superior to that of males on verbal-based items. The extent to which this switch in performance superiority suggests atypical samples of males or females may render these results questionable. On the other hand, our traditional assumptions about gender-determined verbal proficiency may be at fault (Brooks-Gunn & Matthews, 1979).

It is also important to note that these items appeared in experimental units on a pilot basis. Therefore, some of the items may be less satisfactory,

from the perspective of item characteristics (difficulty and discrimination), than is true of items included in the final form of the test. Since the validity of the indices used in the analyses is directly influenced by the nature of the items, the presence of less than satisfactory items may provide another source of variability in the analysis. A replication of the study, based on items with previously established difficulty and discrimination indices, would serve to clarify the impact of cuing that may originate from "generalized" language in the item stem.

The process of phrasing an item precisely, grammatically, yet neutrally is often quite difficult. Nonetheless, it is strongly endorsed. At the very least, the test constructor should be alerted to the possibility of creating cuing situations, however inadvertently, in the item types under investigation. Because this study has found that women perform differentially less well on vowel-consonant items with generalizing cues (and therefore they may be covertly punished for following the cue to an incorrect alternative), the prudent policy would be to adopt neutral language that would remove cue possibilities from the stem.

Finally, given the fact that many of the items were extremely difficult for the students (as evidenced by many low difficulty indices), it is somewhat surprising that more students did not, in fact, pay more attention to the cues in Forms A and C. Using problem-solving strategies on an item that was extremely difficult, a testwise student would theoretically process any and all possible cues in the item. This would have resulted in scores significantly higher on Form A and lower on Form C. Such was not the case. It appears that students are not, in general, paying attention to external information sources in an item. On the other hand, given all the accumulated evidence that general language facility has steadily been declining in this electronic age, the results may not be so surprising after all.

References

Board, C. & Whitney, D. R., The effect of selected poor item-writing practices on test difficulty, reliability and validity, Journal of Educational Measurement, 1972, 9, 3, 225-233.

Brooks-Gunn, J. & Matthews, W. S., He & She: Children's Sex Role Development, Prentice-Hall, Englewood Cliff, New Jersey, 1979.

Dudycha, A. L. & Carpenter, J. B., Effect of item format on item discrimination and difficulty, Journal of Applied Psychology, 1973, 58, 1, 116-121.

Dunn, T. F. & Goldstein, L.G., Test difficulty, validity, and reliability as functions of selected multiple-choice item construction principles, Educational and Psychological Measurement, 1959, 19, 2, 171-179.

Ebel, R. L., Essentials of Educational Measurement, Prentice-Hall, Englewood Cliff, New Jersey, 1972.

Hughes, H. H. & Trimble, W. E., The use of complex alternatives in multiple choice items, Educational and Psychological Measurement, 1965, 25, 1, 117-126

McMorris, R. F., Snyder, G. W., Brown, J. A., & Pruzek, R. M., Effect of violating item construction principles, Journal of Educational Measurement, 1972, 9, 4, 287-295.

Table 1

Subject Description

Experiment	Race		Sex		
Unit #	White	Other	Male	Female	Total
320	201	25	102	152	272
330	197	35	102	153	255
331	204	31	125	132	257
332	203	47	130	137	267

Table 2

Means and Standard Deviations for Percent Chosing Keyed Answer by Sex and by Race for Plural-Singular and Vowel-Consonant Item Types

Plural - Singular Type Items (k = 9)		Form A $\bar{X}$	A s	B $\bar{X}$	B s	C $\bar{X}$	C s	D $\bar{X}$	D s
Sex	M	34.77	11.92	34.76	13.66	33.97	13.52	34.05	13.32
	F	32.19	7.71	31.92	10.99	29.79	12.16	29.5	12.22
Race	W	34.13	10.80	35.10	12.49	32.71	12.32	32.57	13.75
	O	27.09	9.94	22.19	9.83	28.19	14.47	30.90	9.73

Vowel - Consonant Type Items (k = 7)		Form A $\bar{X}$	A s	B $\bar{X}$	B s	C $\bar{X}$	C s	D $\bar{X}$	D s
Sex	M	51.42	11.53	52.32	8.93	54.04	10.88	51.93	9.37
	F	46.10	10.40	45.77	11.14	43.87	13.40	44.40	10.74
Race	W	49.61	10.41	51.60	10.50	49.30	12.30	49.91	15.98
	O	38.21	9.85	36.80	10.44	42.46	15.98	38.41	6.98

Table 3

3 factor (Form, Item, Group) ANOVA F ratios

Sex and Race groups

Plural-Singular and Vowel-Consonant Item Types

	Plural-Singular		Vowel-Consonant	
Source	Sex	Race	Sex	Race
Between				
Items	17.04*	5.6*	8.36*	15.62*
Groups	9.65*	13.95*	6.95*	35.79*
Within				
Form	0.45	0.49	1.20	0.52
Form x Item	6.71*	3.16*	1.48	1.60
Form x Group	0.12	1.70	2.48*	0.80

* $p < .05$

Table 4

Form x Sex Interaction Bonferroni Follow-up

Vowel-Consonant Items

Form	$\bar{X}_i - \bar{X}_j$
A	5.32*
B	6.55*
C	10.17*
D	7.53*

$$CD: \left(\sqrt{\frac{2\ ms_s}{n}}\right).005t_{21}$$

$$(1.58)\ (2.83) = 4.49$$

Figure 1

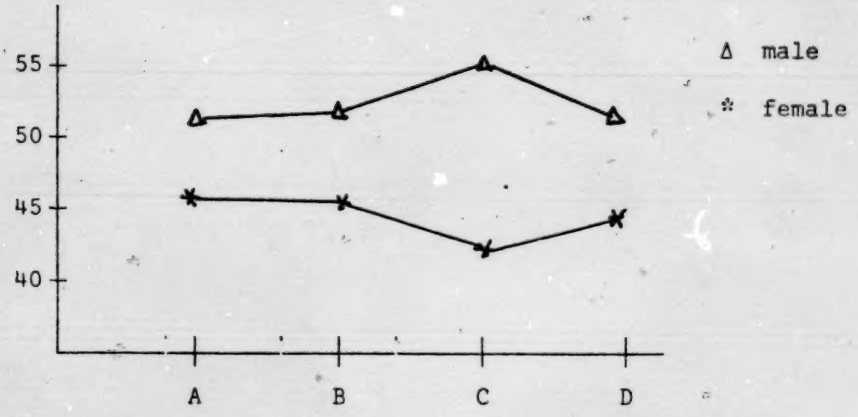

Table 5

Cue Analysis

3 factor ANOVA Summary Tables

Percent chosing keyed response as a

Function of Cue Direction, Cue Status, and Group

Plural-Singular and Vowel-Consonant Item Types

Plural-Singular Item Type (k = 9)

		Group			
		Sex		Race	
Source	df	ms	F	ms	F
Direction	1	43.41	0.30	35.55	0.25
Status	1	0.15	0.01	2.73	0.02
Group	1	221.57	1.52	782.79	5.60*
Direction x Status	1	0.04	0.00	50.35	0.36
Direction x Group	1	11.60	0.08	220.52	1.58
Status x Group	1	0.31	0.02	8.95	0.06
Direction x Status x Group	1	0.00	0.00	81.91	0.59
Residual	64	145.81		142.72	

* $p < .03$

Table 5

Continued

Vowel-Consonant Item Type (k = 7)

		Group			
		Sex		Race	
Source	df	ms	F	ms	F
Direction	1	1.68	0.01	0.83	0.01
Status	1	0.90	0.01	28.29	0.21
Group	1	765.94	6.48**	2207.61	16.39***
Direction x Status	1	4.07	0.03	83.04	0.62
Direction x Group	1	29.57	0.25	21.63	0.16
Status x Group	1	1.75	0.02	103.69	0.77
Direction x Status x Group	1	13.11	0.11	42.88	0.32
Residual	48	118.24		134.68	

** $p < .02$

*** $p < .001$

Table 6

Cue Analysis

Biserial Correlation

Cue-Direction and Cue Status

Plural-Singular and Vowel-Consonant Types

ITEM TYPE	CUE DIRECTION	CUE STATUS	MEAN	STANDARD DEVIATION	k
PLURAL-SINGULAR	leads	present	.2000	.2079	9
		absent	.2544	.1539	9
	misleads	present	.2044	.1272	9
		absent	.1722	.1777	9
VOWEL-CONSONANT	leads	present	.3314	.1447	7
		absent	.3514	.1254	7
	misleads	present	.2817	.1220	7
		absent	.3257	.1350	7

Table 7

Cue Analysis

2 factor ANOVA Summary Tables

biserial correlation as a function of

Cue Direction and Cue Status

Plural-Singular and Vowel-Consonant Types

	Item Type					
	Plural-Singular (k = 9)			Vowel-Consonant (k = 7)		
Source	df	ms	F	df	ms	F
Direction	1	136.11	0.48	1	85.75	0.49
Status	1	11.11	0.04	1	60.04	0.34
Direction x Status	1	169.00	0.59	1	6.04	0.04
Residual	32	286.75		24	174.45	

Table 8

Cue Sensitivity-Direction

χ^2 Goodness of Fit

Plural-Singular Item Types

		Alternatives				
		Inconsistent			Consistent	
		1	2	3	4	5
Percent	Expected	20	20	20	20	20
	Observed	17	13	17	22	28

$$\chi^2_4 = 6.75$$

Table 9

Cue Sensitivity-Direction

χ^2 Goodness of Fit

Vowel-Consonant Item Types

		Alternatives				
		Inconsistent				Consistent
		1	2	3	4	5
Percent	Expected	20	20	20	20	20
	Observed	10	15	10	17	48

$\chi^2_4 = 50.9$ *

* $.95\ \chi^2_4 = 9.49$

Table 10

Cue Sensitivity-Direction

Shift in Consistent Alternative Selection

Plural-Singular Item Types

	Form		Shift
Item	A	B	A - B
1	54	59	-5
2	54	60	-6
3	65	57	8
4	40	43	-3
5	43	63	-20
6	41	35	6
7	61	59	2
8	54	51	3
9	48	46	2
			-13

-1.4 average shift

Table 11

Cue Sensitivity-Direction

Shift in Consistent Alternative Selection

Vowel-Consonant Item Types

	Form		Shift
Item	A	B	A - B
1	49	49	0
2	49	50	-1
3	29	28	-1
4	59	58	1
5	55	55	0
6	12	12	0
7	51	56	-4
			-5

-.71 average shift

Table 12

Cue Sensitivity-Misdirection

χ^2 Goodness of Fit

Plural-Singular Item Types

		Incorrect Alternatives			
		Inconsistent		Consistent	
		1	2	3	4
Percent	Expected	25	25	25	25
	Observed	18	23	33	28

$\chi^2_3 = 5.04$

Table 13

Cue Sensitivity-Direction

χ^2 Goodness of Fit

Vowel-Consonant Item Types

		Incorrect Alternatives			
		Inconsistent			Consistent
		1	2	3	4
Percent	Expected	25	25	25	25
	Observed	25	30	20	16

$$\chi^2_3 = 5.24$$

Table 14

Cue Sensitivity-Misdirection

Shift in Alternative Selection

Plural-Singular Item Types

		Form		Shift
	Item	C	D	C - D
	1	25	34	-9
	2	45	52	-7
	3	62	59	3
Percent	4	37	29	8
Selecting	5	42	43	-1
Alternative	6	34	27	7
	7	38	34	4
	8	34	32	2
	9	44	54	10
				17

average shift 1.89

Table 15

Cue Sensitivity-Misdirection

Shift in Alternative Selection

Vowel-Consonant Item Types

	Item	Form C	Form D	Shift C - D
	1	7	12	-5
	2	4	7	-3
Percent	3	12	14	-2
Selecting	4	5	3	2
Alternative	5	6	10	-4
	6	9	10	-1
	7	11	9	2
				-11

average shift 1.57

Appendix 1

Examples of Item Types*

Singular-Plural Combination Type Items

Form A

Which modern American government agency(ies) was(were) NOT established in the United States Constitution?

I. The Internal Revenue Service
II. The Department of Health, Education, and Welfare
III. The Federal Bureau of Investigation

A. I only
B. II only
C. III only
D. I and II only
(E.) I, II, and III

Form B

Modern American government agencies that were NOT established in the United States Constitution would include the:

I. Internal Revenue Service.
II. Department of Health, Education, and Welfare.
III. Federal Bureau of Investigation.

A. I only
B. II only
C. III only
D. I and II only
(E.) I, II, and III

*Keyed response is circled.

Appendix 1

Examples of Item Types*, Continued

Form C

Which modern American government agency(ies) was(were) NOT established in the United States Constitution?

I. The electoral college
II. The Supreme Court
III. The Internal Revenue Service

A. I only
B. II only
(C.) III only
D. I and III only
E. I, II, and III

Form D

Modern American government agencies that were NOT established in the United States Constitution would include the:

I. electoral college.
II. Supreme Court.
III. Internal Revenue Service.

A. I only
B. II only
(C.) III only
D. I and III only
E. I, II, and III

*Keyed response is circled.

Appendix 1

Examples of Item Types*, Continued

Vowel-Consonant Type Items

Form A

To abolish the Bill of Rights would require a(an):

A. presidential veto.
(B.) amendment to the Constitution.
C. joint resolution of the House and Senate.
D. two-thirds vote of the electoral college.
E. declaration of emergency signed by the secretaries of state and defense.

Form B

To abolish the Bill of Rights would require a:

A. presidential veto.
(B.) constitutional amendment.
C. joint resolution of the House and Senate.
D. two-thirds vote of the electoral college.
E. declaration of emergency signed by the secretaries of state and defense.

Form C

To abolish the Bill of Rights would require a(an):

A. presidential veto.
(B.) constitutional amendment.
C. joint resolution of the House and Senate.
D. two-thirds vote of the electoral college.
E. emergency declaration signed by the secretaries of state and defense.

Form D

To abolish the Bill of Rights would require:

A. a presidential veto.
(B.) a constitutional amendment.
C. a joint resolution of the House and Senate.
D. a two-thirds vote of the electoral college.
E. an emergency declaration signed by the secretaries of state and defense.

*Keyed response is circled.